

This is a published notice on the Find a Tender service: <https://www.find-tender.service.gov.uk/Notice/033130-2024>

Tender

Safeguarded AI: TA1.4 Socio-technical Integration

ADVANCED RESEARCH AND INVENTION AGENCY

F02: Contract notice

Notice identifier: 2024/S 000-033130

Procurement identifier (OCID): ocds-h6vhtk-04ac25

Published 15 October 2024, 12:42pm

Section I: Contracting authority

I.1) Name and addresses

ADVANCED RESEARCH AND INVENTION AGENCY

96 EUSTON ROAD,

LONDON

NW12DB

Email

clarifications@aria.org.uk

Country

United Kingdom

Region code

UKI31 - Camden and City of London

Justification for not providing organisation identifier

Not on any register

Internet address(es)

Main address

<https://www.aria.org.uk>

I.3) Communication

The procurement documents are available for unrestricted and full direct access, free of charge, at

<https://www.aria.org.uk/programme-safeguarded-ai/>

Additional information can be obtained from the above-mentioned address

Tenders or requests to participate must be submitted electronically via

<https://www.aria.org.uk/programme-safeguarded-ai/>

I.4) Type of the contracting authority

Body governed by public law

I.5) Main activity

General public services

Section II: Object

II.1) Scope of the procurement

II.1.1) Title

Safeguarded AI: TA1.4 Socio-technical Integration

II.1.2) Main CPV code

- 73110000 - Research services

II.1.3) Type of contract

Services

II.1.4) Short description

ARIA is an R&D funding agency built to unlock scientific and technological breakthroughs that benefit everyone. We empower scientists and engineers to pursue research at the edge of what is technologically or scientifically possible. We will reach across disciplines, sectors and institutions to shape, fund and manage projects across the R&D ecosystem, from startups to universities, to break down silos and discover new pathways.

We are looking for proposals for our Safeguarded AI: TA1.4 Sociotechnical Integration For more info see here <https://www.aria.org.uk/programme-safeguarded-ai/>

II.1.6) Information about lots

This contract is divided into lots: No

II.2) Description

II.2.3) Place of performance

NUTS codes

- UK - United Kingdom

II.2.4) Description of the procurement

Why this programme: as AI becomes more capable, it has the potential to power scientific breakthroughs, enhance global prosperity, and safeguard us from disasters. But only if it's deployed wisely.

Current techniques working to mitigate the risk of advanced AI systems have serious

limitations, and can't be relied upon empirically to ensure safety. To date, very little R&D effort has gone into approaches that provide quantitative safety guarantees for AI systems, because they're considered impossible or impractical.

What we're shooting for: by combining scientific world models and mathematical proofs we will aim to construct a 'gatekeeper', an AI system tasked with understanding and reducing the risks of other AI agents.

In doing so we'll develop quantitative safety guarantees for AI in the way we have come to expect for nuclear power and passenger aviation.

Our goal: to usher in a new era for AI safety, allowing us to unlock the full economic and social benefits of advanced AI systems while minimising risks.

The third solicitation for this programme is focused on TA1.4 Socio-technical Integration. Backed by £3.4m, we're looking to support teams from the economic, social, legal and political sciences to consider the sound socio-technical integration of Safeguarded AI systems.

This solicitation seeks R&D Creators - individuals and teams that ARIA will fund - to work on problems that are plausibly critical to ensuring that the technologies developed a part of the programme will be used in the best interest of humanity at large, and that they are designed in a way that enables their governability through representative processes of collective deliberation and decision-making.

A few examples of the open problems we're looking for people to work on:

- Qualitative deliberation facilitation: What tools or processes best enable representative input, collective deliberation and decision-making about safety specifications, acceptable risk thresholds, or success conditions for a given application domain? We hope to integrate these into the Safeguarded AI scaffolding.
- Quantitative bargaining solutions: What social choice mechanisms or quantitative bargaining solutions could best navigate irreconcilable differences in stakeholders' goals, risk tolerances, and preferences, in order for Safeguarded AI systems to serve a multi-stakeholder notion of public good?
- Governability tools for society: How can we ensure that Safeguarded AI systems are governed in societally beneficial and legitimate ways?
- Governability tools for R&D organisations: Organisations developing Safeguarded AI capabilities have the potential to create significant externalities - both risks and benefits. What set of decision-making and governance mechanisms are best to ensure that entities developing or deploying Safeguarded AI capabilities have and maintain these externalities

as appropriately major factors in their decision-making?

We are also open to applications proposing other lines of work which illuminate critical socio-technical dimensions of Safeguarded AI systems, if they propose solutions to increase assurance that these systems will reliably be developed and deployed in service of humanity at large.

II.2.5) Award criteria

Price is not the only award criterion and all criteria are stated only in the procurement documents

II.2.7) Duration of the contract, framework agreement or dynamic purchasing system

Duration in months

18

This contract is subject to renewal

No

II.2.10) Information about variants

Variants will be accepted: No

II.2.11) Information about options

Options: Yes

Description of options

Additional funding, scope and duration could be added to any contracts awarded.

Section IV. Procedure

IV.1) Description

IV.1.1) Type of procedure

Competitive procedure with negotiation

IV.1.8) Information about the Government Procurement Agreement (GPA)

The procurement is covered by the Government Procurement Agreement: No

IV.2) Administrative information

IV.2.2) Time limit for receipt of tenders or requests to participate

Date

2 January 2025

Local time

12:00pm

IV.2.4) Languages in which tenders or requests to participate may be submitted

English

Section VI. Complementary information

VI.1) Information about recurrence

This is a recurrent procurement: No

VI.3) Additional information

Detailed timelines can be found in the programme call information on ARIAs website:
<https://www.aria.org.uk/programme-safeguarded-ai/>

The deadline for submission of a full proposal is 02 January 2025 (12:00 GMT).

The total funding value is the estimated budget available. We expect to fund multiple applicants.

Funding is anticipated to be award via both contracts and grants. For information on how we fund <https://www.aria.org.uk/faqs-funding/>

VI.4) Procedures for review

VI.4.1) Review body

See the ARIA Act 2022

London

Country

United Kingdom